AD-A283 221

NTATION PAGE

Form Approved
OMB No. 0704-0188

sted to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this burden, to Washington Headquarters Services, Directorate for information Operations and Reports, 1215 Jefferson Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

ORT DATE	3. REPORT TYPE AND DATES COVERED

4. TITLE AND SUBTITLE
The Criteria-Based Content Analysis and its Utility in Distinguishing Between Truthful and Fabricated Criminal Allegations.

5. FUNDING NUMBERS

6. AUTHOR(S)
Ruby & Brigham

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Florida State University

8. PERFORMING ORGANIZATION REPORT NUMBER
94-105

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
DEPARTMENT OF THE AIR FORCE
AFIT/CI
2950 P STREET
WRIGHT-PATTERSON AFB OH 45433-7765

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release IAW 190-1
Distribution Unlimited
MICHEAL M. BRICKER, SMSgt, USAF
Cheif Administration

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

14. SUBJECT TERMS			15. NUMBER OF PAGES 606
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT

NSN 7540-01-280-5500

Standard Form 298

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to ***stay within the lines*** to meet ***optical scanning requirements***.

Block 1. Agency Use Only *(Leave blank)*.

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C	- Contract	**PR**	- Project
G	- Grant	**TA**	- Task
PE	- Program Element	**WU**	- Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. *(If known)*

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of...; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

- **DOD** - See DoDD 5230.24, "Distribution Statements on Technical Documents."
- **DOE** - See authorities.
- **NASA** - See Handbook NHB 2200.2.
- **NTIS** - Leave blank.

Block 12b. Distribution Code.

- **DOD** - Leave blank.
- **DOE** - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
- **NASA** - Leave blank.
- **NTIS** - Leave blank.

Block 13. Abstract. Include a brief *(Maximum 200 words)* factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code *(NTIS only)*.

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.

94-105

The Criteria-Based Content Analysis and its Utility in Distinguishing Between Truthful and Fabricated Criminal Allegations: A Critical Review

C. L. Ruby and John C. Brigham

Florida State University

ABSTRACT

This review examined the utility of the Criteria-Based Content Analysis as a method to distinguish truthful from fabricated criminal allegations. It appears that the U.S. justice system would accept the CBCA as a viable method for truth detection if empirical support could be obtained. Although the research results demonstrate some utility, the studies have either not been experimental in nature and fall prey to selection bias, or have fallen short of addressing several other important issues. This review suggests some areas that should be addressed in the future.

During the past several years, allegations of child sexual abuse have steadily increased (Wakefield & Underwager, 1988, 1991). The number of child maltreatment reports climbed tremendously from 669,000 in 1976 to 1.9 million in 1985. Of the 1985 figure, 200,000 reports were for sexual abuse (Bulkley, 1989). Frequently, these allegations are made by the alleged victim, whose testimony is the only evidence upon which prosecution or intervention has been based (Coolbear, 1992; Yuille, 1988a; Undeutsch, 1989). To investigate the extent to which children in these situations should be believed, many researchers and professional authorities have investigated developmental aspects of child witness testimony such as susceptibility to suggestions, memory accuracy, and behavioral indicators of sexual trauma. Nevertheless, these research efforts have been criticized in terms of their failure to address the intentions of the child, namely, whether he/she is being truthful (Raskin & Esplin, 1991a). The Criteria-Based Content Analysis (CBCA) is an interview technique that addresses precisely this issue. It is used by some legal professionals as an aid in making judgments about a child's veracity.

The purpose of this critical review is to assess the utility of the CBCA in distinguishing between a truthful and deceptive allegation. This review begins with a historical perspective on the way that professionals have responded to the issue of child sexual abuse allegations and whether there is a role for the CBCA to play in the criminal justice system. That is followed by a brief look at current attempts to detect deception and truth-

telling. Then, we describe the CBCA and its developers' rationale, plus their attempts to empirically assess its utility. This critical review concludes by addressing shortcomings of those studies and some ideas for future research.

A Historical Perspective

In a historical review of the suggestibility of child witnesses, Ceci and Bruck (1993) noted that the prevailing legal stance in the United States from colonial times to the present has been one of skepticism about the testimony of child witnesses. They cited Varendonck (1911) as an example of the early experimental support for this skeptical posture. In assistance to the defense in a murder trail in which a key eyewitness was a child, Varendonck conducted a study in which he asked children to describe a man who allegedly was in their playground. Although in reality the man was never there, the children reported detailed descriptions of him. Varendonck's conclusions about the treatment of child witnesses were that, "we cannot set the least value in their declarations."

In addition to psychologists, legal authorities and scholars have also been skeptical of the credibility of children. In fact, prior to the 1970's, the legal system seemed unwilling to recognize that allegations of child sexual abuse were credible at all. For instance, in his seminal Treatise on Evidence, Wigmore (1940) insisted that all allegations of sexual abuse should be initially assumed false. He used the prevailing scientific literature of his time to support this contention. Wigmore relied heavily on Freud's

ideas of young girls' sexual fantasy towards their fathers as the basis of any allegations they may make.

The skeptical nature of Wigmore's writings established precedence for subsequent jurisprudence in dealing with the admissability of evidence encountered in the courts today (Bienen, 1983). For example, the Oregon Supreme Court in State v. Yates (1965) said the lower court should have instructed the jury of the dangerousness of convicting a defendant based only on the uncorroborated testimony of the child victim (Bienen, 1983). Also, a Florida appeals court ordered a retrial in Hawkins v. State (1976) because the trial court refused to allow Hawkins to obtain psychiatric evidence that could refute the credibility of the alleged child sexual abuse victim whose testimony was the only evidence against Hawkins. These two cases are examples of the characteristic reservation the U.S. courts had into the 1970's. The courts were cautious about the testimony of sexual abuse victims and often found reasons to disbelieve or even discredit the allegations (Bienen, 1983).

According to Yuille (1988a), the 1970's saw an increase in systematic empirical research on the credibility of child witnesses. Whether as a consequence of the research, or as an antecedent, there was a corresponding increase in professionals' confidence in children's allegations. In a review of the current posture toward child sexual abuse allegations, Coolbear (1991) concluded that most professionals generally accept the view that the majority of children do not lie about sexual abuse allegations.

Faller (1984) had earlier said, "we know that children do not make up stories asserting they have been sexually molested." She further contended that researchers and clinicians working in the field of sexual abuse are in agreement that false allegations of child sexual abuse are extremely rare. But, Coolbear (1991) argued there are indeed factors that predispose false allegations to be made (e.g., custody disputes, anger toward a parent, prodding by parent, and method of interviewing).

The uncritical acceptance of a child's allegation such as that displayed by Faller (1984) can be a serious problem. The alleged child victim is many times the only witness, and his or her testimony has resulted in the conviction and imprisonment of alleged perpetrators. Because of the precarious situation for those accused of child sexual abuse, many researchers have tried to assess the prevalence of false sexual abuse allegations. According to Quinn (1988), the estimate of false allegations reported by numerous studies have varied widely, ranging from as low as 3% to as high as 75%. Clancy and Coleman (1990) asserted that although the experts may disagree as to the prevalence of false accusations, they are happening at a rate far in excess of what our society can afford. They believe false allegations are primarily the result of an working alliance between law enforcement and mental health agencies. In their view, the traditional law enforcement focus on objectively ferreting out the facts has been contaminated by the infusion of a therapeutic, welfare-promoting approach of the mental health profession that may encourage false allegations.

This apparent trend towards more confidence in children's allegations seems to be limited to law enforcement officials and other professionals who are the first to respond to sexual abuse allegations. On the other hand, the U.S. courts continue to struggle with the question of child victims' believability and whether methods can be used to assess the child's credibility. In referring to the conflict between wanting to admit testimony of a child sexual abuse victim on one hand, and the need for reliable testimony on the other, Rozell (1985) said that the courts, "should resolve this conflict by deferring to sound principles of human behavior as reflected in behavioral data." She emphasized that it is important to rely on scientific methods to establish the child's credibility. Nevertheless, in reviewing the methods used by courts to determine credibility, Rozell (1985) noted that Germany and Austria had been the only countries to employ the CBCA method to analyze a child's veracity. Additionally, a more up-to-date search of U.S. state and federal court proceedings through 1992 failed to identify any use of the CBCA to distinguish false from truthful allegations of child sexual abuse.

Even though the U.S. courts continue to struggle with how to establish a child's believability, they have not sufficiently addressed the issue of the <u>veracity</u> of the child's allegation. Instead, they have focused on whether or not to admit expert psychological testimony about the child's <u>credibility</u>, or reputation for truthfulness. In other words, the focus has been on the child (credibility) instead on the allegation (veracity). The

closest the U.S. courts have come to the issue of an allegation's veracity has been with respect to the "behavioral syndrome" exhibited by the child (Baker, 1990; Coleman & Clancy, 1990; Nicholas, 1988). Examples of these behaviors are delayed disclosure of abuse, emotional response when talking about the abuse, and sexual responses to projective tests. They are thought to be indicative of the traumatic experience of sexual abuse (see Green, 1986; Nurcombe, 1986; & Sink, 1988 for examples of these behavioral syndromes). The focus has been on whether the usefulness of these behaviors as indicators of sexual abuse can be scientifically supported and, therefore, give credence to the allegation. But, as will be explained shortly, non-verbal (i.e., behavioral) cues may not be an accurate indicator of truthfulness. Even though the behavioral syndrome might fall short of providing adequate evidence of the allegation's veracity, the U.S. courts have usually not been agreeable in allowing other expert testimony (e.g., CBCA results) to specifically address the veracity of a particular child's statement (Undeutsch, 1989).

For example, in Commonwealth v. Seese (1986), the Pennsylvania Supreme Court ruled that experts may not address the veracity of a child sexual abuse victim based only on the expert's professional experience with that class of witnesses. The court said that jurors have the responsibility to judge the truthfulness of a particular witness and that allowing an expert to do so would invade the province of the jury. The court suggested that a more moderate approach would be to allow expert testimony of child

sexual abuse victim symptomology (i.e., behavioral syndrome) but not opinions as to the particular allegation's veracity.

In a similar situation in U.S. v. Azure (1986), a U.S. Court of Appeals said that "putting an impressively qualified expert's stamp of truthfulness on a [child] witness' story goes too far." And with respect to the scientific validity of judging a particular child's truthfulness, the court concluded that "no reliable test for truthfulness exists."

A more liberal application of expert testimony that addresses the veracity of abuse allegations is found in State v. Kim (1982). Here, the Hawaiian Supreme Court ruled that an expert's testimony concerning a child victim's veracity was proper because it was of value to the jury and far outweighed any prejudicial effects it might have had on the defendant. But, in this case, the expert's testimony consisted of specific behaviors exhibited by the victim that, according to the expert witness, indicated a truthful allegation. These indicators appeared to be a mixture of "behavioral syndrome" behaviors similar to those mentioned in Commonwealth v. Seese, and particulars about the victim's verbal content (e.g., consistency of the allegation) as addressed by the CBCA. Furthermore, the Kim court said that expert testimony concerning a victim's veracity should be based on a sound foundation, a reliable system of analysis, a sound methodology, and precise inferences.

Although most states follow the position taken in Commonwealth v. Seese and do not allow experts to testify as to the child

victim's veracity, some follow State v. Kim (e.g., State v. French, Montana 1988; State v. Geyman, Montana 1986; State v. Myers, Minnesota 1984; State v. Timperio, Ohio 1987). In reviewing these cases, Baker (1990) concluded that experts should be allowed to testify as to the veracity of a particular child witness. As the court suggested in Kim, the debate over whether to allow expert testimony about the victim's veracity might be better resolved if a scientifically validated method can be used to indicate veracity. If empirical evidence can be used to demonstrate the diagnostic utility of the CBCA, the courts might be more inclined to permit its use to assist the jury in determining the veracity of sexual abuse allegations.

Throughout this discussion of the criminal justice system's reaction to allegations of child sexual abuse, the distinction has been made between a child's credibility and the veracity of his or her allegation. Köhnken (1989a) used this distinction in outlining two approaches that have typically been used to assess a child witness' testimony. The first approach employs a social judgement or an impression formation process to arrive at an assessment of the child's credibility. This process involves subjective evaluations of the child's demeanor, personality, and reputation, and then a global assessment of the child's credibility.

The second approach is what Köhnken (1989a) termed a "psychodiagnostic decision-making task." This is the method used by psychological experts in an attempt to provide empirical evidence to support or to refute the veracity of the child's

statement, not his or her overall credibility. Undeutsch (1989) likewise emphasized the need to address the veracity of the particular statement and not the credibility of the particular child witness. The second approach would be endorsed by the court in Kim. This is also the approach pursued when applying the CBCA.

With respect to this distinction, Köhnken and Steller (1988) evaluated the German "inquisitorial" legal system with the "adversarial" legal system of Great Britain, Canada, and the United States. In the inquisitorial system, the main aim is fact-finding. In the adversarial system, the main aim is to support one's client (government or defendant). A fact-finding system emphasizes objectivity and uses the psychodiagnostic decision-making approach to determine the veracity of a statement. In 1954, the Supreme Court of the Federal Republic of Germany mandated that expert psychologists or psychiatrists provide an assessment of the veracity of a child witness if the child's testimony is the main evidence against the defendant. This cleared the way for the widespread use of the CBCA in that country (Raskin & Yuille, 1989).

On the other hand, an adversarial system tends to encourage global references to either the prosecution's or defense's general reputation or credibility and also to encourage investigators to prove their position rather than seek out facts consisting of incriminating and exculpatory evidence (see Raskin & Yuille, 1989; Scott, 1989). Hence, the adversarial nature of the U.S. court system makes it easier to employ global credibility assessments rather than psychodiagnostic procedures such as the CBCA. This is

reflected in the above-cited court decisions, the majority of which shy away from allowing testimony about the victim's veracity but which allow testimony about his or her credibility (propensity or reputation for truthfulness).

Other researchers have shown that professionals in an adversarial system fail to use psychodiagnostic techniques to establish the veracity of an allegation. Moreover, this failure may have led to false allegations of child sexual abuse (Coolbear, 1991; Yuille, 1988a). In a study that compared techniques used by legal professionals with those used by human services professionals in Canada, Coolbear (1991) found that the interviewers' professional roles dictated the method they used to interview child victims, rather than their mission of fact-finding. More specifically, it appeared that the interviews of child sexual abuse victims were conducted in a manner that was intended to confirm each agency's assumptions about the case (adversarial) instead of determining the facts surrounding the case (fact-finding). In a similar fashion, Yuille (1988a) pointed out that police interviews tend to be rigid, structured, and have a set agenda. The primary goal of these interviews appeared to be an attempt to prove a particular assumption rather than to investigate possibilities. These two investigators suggest that legal and human services professionals' adversarial approaches interfered with the task of determining the veracity of sexual abuse allegations and may in fact have elicited false allegations.

The foregoing review of the criminal justice system's attempts

at distinguishing between truthful and deceptive allegations of child sexual abuse makes it clear that a technique such as the CBCA, if it provided valid information, would be very useful in addressing the veracity of child sexual abuse allegations. The courts' apparent reluctance to allow expert testimony concerning the child sexual abuse victim's veracity may be due to the lack of a convincing, scientifically sound, method to do so. The lack of such a reliable technique along with the adversarial nature of the U.S. criminal justice system seems to have forced U.S. professionals into the role of advocate rather than fact-finder.

Current Efforts in Detecting Deception

In continuing to determine whether the CBCA can be useful in the criminal justice system, we must also assess the extent to which present efforts in detecting deception (and truthfulness) are successful. People generally judge themselves very good at detecting deception in others (Köhnken, 1989a). They also judge the authorities to be good at detecting deception. For instance, in a survey of 101 jurors, 98% of them indicated that they believed mental health professionals could tell whether or not a child was actually sexually abused. Even in cases where there is no evidence, 82% thought those professionals could do so (Corder & Whiteside, 1988).

But, research has shown that neither laypersons nor professionals are much better than chance in detecting deception (Ekman & O'Sullivan, 1991; DePaulo, Stone, & Lassiter, 1985; Zuckerman, DePaulo, & Rosenthal, 1981). In a survey of the non-

verbal indicators law enforcement officers use to detect deception of witnesses, Ruby and Brigham (1992) found that despite the professional training, law enforcement officers use many of the same cues in attempting to detect deception as do college students. Professionals who are usually thought to be trained in detecting deception (e.g., police officers, immigration and customs agents, psychologists, etc.) have been shown not to be much better than laypersons (DePaulo & Pfeifer, 1986; Köhnken, 1987; Kraut & Poe, 1980). Eckman and O'Sullivan (1991) found that of police robbery investigators, federal polygraphers, psychiatrists, and Secret Service agents, only the latter were significantly better than chance at detecting deception of videotaped presentations.

Köhnken (1989a) attributes this erroneous over-estimation of the ability to detect deception to the cognitive salience of successful attempts, and the ignoring of unsuccessful attempts. This leaves the person susceptible to the effects of "availability heuristics" (Tversky & Kahneman, 1974) and "confirmatory bias" (Snyder & Swann, 1978), in which those salient instances are used to confirm the person's assumption that he or she is successful in detecting deception. Another common error made in trying to detect deception is that people generally think that non-verbal behavior is more valuable in determining whether someone is lying or telling the truth, so they rely on non-verbal behavior more than verbal behavior. Research that compares the extent to which verbal and non-verbal behavior correlate with truthful or deceptive statements suggests that a person's non-verbal behavior may not be a

consistently significant correlate of deception (Eckman & O'Sullivan, 1991; Köhnken, 1989a). In contrasting the use of non-verbal and verbal cues to detect lying, DePaulo, Zuckerman, and Rosenthal (1980) distinguished between "leakage" and "deception." They asserted that a speaker "leaks" non-verbal cues that reveal the speaker's true, concealed emotions, not necessarily the emotions of lying. On the other hand, verbal cues are more revealing of lying, in other words, concealing information. Accordingly, the importance of a child's verbal content, as opposed to non-verbal accompanying behavior, has been the focus of the CBCA method in providing evidence of an allegation's veracity.

It appears, then, that neither social services professionals nor laypersons are accurate at distinguishing between true and deceptive information, although they believe they are. Their misperceptions may be due to cognitive biases and a reliance on non-verbal cues to deception. This makes it clear that the CBCA can play an important role in the criminal justice system.

Criteria-Based Content Analysis

In an attempt to provide a systematic method that assesses the veracity of a child's specific allegation of sexual abuse, rather than the child's overall credibility or reputation for truthfulness, investigators in Europe focused on the verbal content of the child's statement (Undeutsch, 1982). Verbal content criteria were established that were purported to distinguish truthful content from false content.

The CBCA is actually only one portion of a more comprehensive

interview technique called the Statement Validity Assessment (SVA). As described by Raskin and Esplin (1991a), the SVA consists of three parts: (1) a structured interview of the child witness, (2) the CBCA: a systematic analysis of the verbal content of the child's statements, and (3) the application of the Statement Validity Checklist, which addresses non-verbal characteristics of the interview process.

The structured interview portion consists of an extensive interview with the alleged child victim without the use of leading questions. The purpose of this portion of the SVA is to create rapport and assess the child's cognitive, behavioral, and social skills. The interviewer first attempts to determine what the child knows by using a free recall style of interview. The interview starts out with asking the child to relate the incident as fully as he/she remembers. Only after that are more specific open-ended questions asked to clarify inconsistencies or gaps in the information.

The second portion of the SVA consists of the CBCA. In this portion, 18 criteria (See Table 1) are applied to the content of the child's statement and provide a probability estimate of the statement's veracity. The presence of a criterion is an indication that the child is telling the truth. During this portion, it is important to consider the child's age, experience, and skill level when applying the criteria (e.g., younger children's verbal statements may contain less detail, which is one of the CBCA criteria).

Insert Table 1 About Here

The last portion of the SVA consists of applying the Statement Validity Checklist which contains statement-related factors that assess the validity of several other characteristics of the interview process (See Table 2). These characteristics include, for instance, the child's psychological status and things about the interview that may have influenced the content. Based on an integration of the results of these three parts of the SVA, an overall evaluation is made of the statement's veracity.

Insert Table 2 About Here

The SVA is the end product of four decades of collaboration between researchers and practitioners from the United States, Canada, Sweden, and Germany. It evolved from the Statement Reality Analysis (SRA) technique which had been used in Europe since the 1950's. The SRA contains many of the same criteria and checklist items found in the present-day SVA, but is less detailed in its analysis of a child's statement (See Table 3).

Insert Table 3 About Here

The first development of the SRA can be attributed to Undeutsch (1967) (cited in Steller, 1989). At about the same time,

the Swedish investigator Trankell developed a similar technique (Trankell, 1957, 1963) (cited in Yuille, 1988a). Trankell (1972) distinguished between his technique and that of Undeutsch's and later researchers in that his was based more on a holistic or intuitive approach rather than on the reductionistic analysis of specific content criteria of a statement, namely, that process attributed to the CBCA. Besides Undeutsch's and Trankell's systems, other European descriptions of similar statement analyses can be found in Arntzen (1970) (cited in Köhnken & Steller, 1988), Szewczyk (1973) (cited in Köhnken & Steller, 1988), and Dettenborn, Froehlich, and Szewczyk (1984) (cited in Köhnken & Steller, 1988).

Later, North American researchers Yuille and Raskin working in concert with European researchers Steller and Köhnken, modified the SRA and changed its name from the Statement Reality Analysis to the Statement Validity Assessment to better define what they saw was the end result of the process, namely, the validity of a child's statement (Yuille, 1988a). Nonetheless, the SRA is still advocated by Undeutsch (1989) in corroborating allegations of child sexual abuse.

The CBCA has been proclaimed as the core and most important part of the SVA analysis (Raskin & Yuille, 1989) and the central topic of this critical review. The CBCA deals solely with the <u>verbal content</u> of the witness' statement or allegation. Arntzen provided the first classification of the specific CBCA criteria that could be applied to a child's statement (Steller & Köhnken, 1989). But, Köhnken (1982) (cited in Steller & Köhnken, 1989)

criticized Arntzen's criteria as unsystematic and unconvincing. Steller and Köhnken (1989) integrated the criteria provided by Arntzen and previous investigators to arrive at 19 content criteria which they contended were more systematically organized and precise. Raskin and Esplin (1991a) recently argued for the elimination of the 19th criterion from the CBCA and placed it in the third portion of the SVA. This 19th criterion can been seen as the 13th item in the Statement Validity Checklist (Table 2). Raskin and Esplin felt that this item was more characteristic of investigative questions relative to the statement as a whole, rather than verbal content criteria addressed in the CBCA.

The criteria of the CBCA have most recently been divided into three categories: general characteristics, specific contents, and motivation-related contents (Raskin & Esplin, 1991a) (See Table 1). Steller and Köhnken (1989) earlier organized the criteria into five categories. Criteria 8 through 13 were separated from the Specific Contents category and comprised a third one called Peculiarities of the Content. In addition, as mentioned above, a 19th criterion was the sole member of a fifth category called Offense-Specific Elements. Within each of the three current categories, there are specific criteria that the interviewer determines are either present or absent. The presence of one of these specific content characteristics indicates a truthful statement while the absence of them indicates nothing (Steller, 1989; Yuille, 1988a).

Criteria 1 through 3 fall within the <u>general characteristics</u> category, and are applied to the statement as a whole with regard

to its logical structure, unstructured production, and quantity of details. Having a **logical structure** means the statement's details are integrated into a whole and have no contradictions or logical inconsistencies. They independently describe the same series of events. To have a logical structure, the statement does not necessarily have to be chronological or complete in describing the incident in question. In fact, an **unstructured production** in a statement also signifies a truthful account. This characterizes the statement as discontinuous and fragmented. Fabricated stories are usually sequential and chronological with clear attempts to demonstrate causal connections. The last item within the general characteristics category is **quantity of details.** If a statement has a high number of non-repeated factual details, its veracity is indicated.

The second category contains criteria 4 through 13 and deals with specific contents of the statement. These criteria are the amount of contextual embedding, descriptions of interactions, reproduction of speech, unexpected complications during the incident, unusual details, superfluous details, accurately reported details which are misunderstood, related external associations, accounts of the victim's subjective experiences, and accounts of the accused's mental state. If an incident is related within the context of spatial or temporal events existing at the time of the event, it is said to have **contextual emb dding.** A person who relates description of **interactions** is recounting interpersonal interactions, either fluently or awkwardly, accurately or

inaccurately. Interactions must consist of at least three "elements," that is the victim and accused must have three specific action/reaction exchanges in order for one interaction to occur. **A reproduction of speech** means that the person being interviewed is reporting a verbatim dialogue, especially when the words used are atypical for a child, when reasons are presented, or when attitudes are revealed. Any **unexpected complications** reported during the incident indicate a truthful statement. An example would be if a child mentions that in the middle of some sexual act, the accused had to stop to answer the telephone. A child's statement that relates **unusual details** or **superfluous details** suggests credibility. Also, if a child incorrectly interprets a correctly described detail, he or she has **accurately reported details misunderstood.** Such details might be when a child reports that a male accused urinated when describing ejaculation. If the interviewee describes other sexual events or conversations which happen at the same time as the relevant incident, but external to it, he/she is describing **related external associations.** A related external association might consist of the accused discussing his/her previous sexual encounters with the victim, or the victim's previous sexual experiences with others. The reporting of emotions and cognitions occurring to the child during the event are examples of the child's **subjective experience.** Lastly, if a child makes an **attribution of the accused's mental state,** this suggests that he or she is being truthful.

The third category deals with motivation-related contents.

These refer to the motivation of the child making the allegation and include spontaneous corrections or additions, admissions of a lack of memory or knowledge, doubts about his or her own testimony, self-deprecation, and pardoning the accused. These characteristics demonstrate behavior that a deceptive person rarely shows because they tend to minimize the allegation. Making **spontaneous corrections or additions** show that the victim has made a mistake and is willing to correct it. The key here is that the correction be spontaneous, not corrected after pondering over the mistake or after direct questioning by the interviewer. A child who **admits a lack of memory or knowledge, raises doubts about his or her own testimony,** or engages in **self-deprecation** is similarly demonstrating doubt as to an ability to recall accurately or questioning his/her own worth. Deceptive people would rarely do so if they are concerned about appearing truthful. Finally, a child who **pardons the accused** goes against what would normally be expected of a deceptive person because this excuses the alleged perpetrator's actions.

As already mentioned, another category that was part of the CBCA until recently was offense-related elements. This contained the single criterion of **details characteristic of the offense** that the child reports. These include details that are counter to the common sense view, but accurate characteristics of the crime of child sexual abuse usually only known by the experts. To apply this criterion, the interviewer must have enough expertise in the offense area in order to realize when details are congruent with

the crime of child sexual abuse or against the common sense view. But, as pointed out by Horowitz (1991), it is difficult to determine just what constitutes a common sense view of offenses. This is especially true as society becomes more educated about sexual issues and more emphasis is being placed on abuse of children. Accordingly, Raskin and Esplin (1991a) recently moved this criterion to the third portion of the SVA which addresses the validity of the statement as a whole.

The CBCA criteria are based on Undeutsch's assumption that memory of an actual experience differs in verbal quality and content from statements which are invented, or in other words, not actually experienced by the person making the statement. People who are making up a story are unlikely to speak as if they are emotionally and cognitively re-experiencing the event. For example, actual re-experiencing of the event would be manifested by reporting actual speech (criterion 6), reporting feelings experienced at the time of the event (criterion 12), relating extraneous intrusions into the event (criterion 7, 11), or reporting an event that they did not understand (criterion 10). Furthermore, bogus victims would not likely cast doubt as to their memory (criterion 15, 16) or forgive their attackers (criterion 18). If they were trying to deceive, they probably would attempt to strengthen their accusation as much as possible (unless, however, they are aware of the CBCA rationale). However, it must be remembered that falsified stories may contain some of the criteria if the alleged victim has had actual prior experience with

sexual matters, either an early maturer or the child has been abused by some person, not necessarily the accused (Raskin & Esplin, 1991a).

The CBCA was based on clinical experience and addresses factors that were thought by the above researchers to differentiate a truthful from a deceptive child victim's statement. Some studies have shown that professionals involved with allegations of child sexual abuse have in fact intuitively relied on some CBCA criteria. In looking at the opinions of 51 legal and human service professionals, Coolbear (1991) showed that some of these professionals unwittingly used some of the CBCA criteria, although not many of them. In another study which asked law enforcement officers what they used as indicators of truthful/deceptive statements, these professionals claimed they considered consistency of a story (i.e., criterion 1 -- logical structure) as a cue to truthfulness (Ruby & Brigham, 1992).

Köhnken (1989b) discussed two major processes involved in truth-telling and deceiving that are expected to influence the verbal content of a speaker's message and which may provide theoretical support for some of the CBCA criteria. First, in an attempt to control what he or she says, a deceiver's message content may appear planned, not spontaneous, or otherwise not natural. The second fundamental process that may affect verbal content involves cognitive processes. Social cognitive schema theory would predict that fabricated stories would be constructed from a person's previously developed schemas about the central

topic of the fabricated story. Therefore, unusual details, or accounts of details which were misunderstood by a speaker, for example, would be expected from a truthful speaker since such details would have been only recently experienced by the speaker and not incorporated into his or her personal schemas. In other words, a truth-teller's verbal content, in comparison to that of a deceiver's, should consist of more parts that are not related to their personal schemas. On the other hand, the deceiver, because he or she is relying on previously developed personal schemas to invent the story, should communicate less unusual details or misunderstood details.

Köhnken (1982) asserted that the various criteria have different values for assessing the veracity of a statement. For example, he felt that offense-related elements probably have more weight than contextual embedding or descriptions of interactions. According to Steller (1989), criteria one and three serve an extraordinarily important function relative to the remaining criteria. This makes it tenuous to establish a specific number of criteria that must be present in order to assume veracity. Likewise, Littmann and Szewczyk (1983) (cited in Steller and Köhnken, 1989) and Szewczyk and Littmann (1982) (cited in Steller and Köhnken, 1989) concluded that the process of applying the CBCA is interactive; that is, the unique configuration of criteria in a statement and the differing values or weights of criteria are what lead to a truthful or deceptive conclusion. The process is not additive in that a certain number of criteria present can be used

to indicate truthfulness. However, Arntzen (1970) noted that in Germany, there is a rule of thumb that at least three criteria must be present to make a truthful judgement. Also, Landry and Brigham (1992) used the presence of five or more criteria as an indicator of truthfulness in training their subjects. Although points are assigned for the presence of criteria (usually 0 = absent, 1 = present, and 2 = strongly present), the current level of development precludes the establishment of specific weights of criteria or cut-off scores (Steller, 1989).

Steller (1989) identified three rules for applying the CBCA criteria. First, repetitions of a criterion in different passages of a statement do not increase the rating of that criterion. Second, one passage in a statement can fulfill more than one criterion. Last, the contents of the statement are considered only if they are related to the alleged incident.

In summary, the CBCA was a collaborative effort between researchers in Europe and North America, developed on the basis of intuitive and clinical experiences. There is also some theoretical rationale for its development, but empirical evidence is needed to systematically provide support for its utility. That will be addressed in the next section.

Attempts to Empirically Validate the CBCA

Research Paradigms

Steller and Köhnken (1989) concluded that the empirical basis of the CBCA to that date was unsatisfactory. In discussing the ways to validate the CBCA for forensic purposes, Steller (1989)

placed higher value on field studies than on retrospective studies of adjudicated cases or on simulation (experimental) studies. In apparent contradiction to this position though, Steller and Köhnken (1989) were particularly disparaging of field studies, and attributed the lack of empirical foundation to the predominance of field studies in the literature. Their specific concern about the following field studies was that they used highly selective and unambiguous cases of alleged child sexual abuse and relied too heavily on the results of external criteria (e.g., confessions, convictions, recanted testimony) as diagnostic of truthfulness of a child's statement. Steller and Köhnken (1989) therefore recommended increased experimental manipulation of factors affecting the validity of the CBCA. Others have echoed this criticism of field studies and encouraged more development of an empirical, as well as theoretical, foundation (Horowitz, 1991; Wells & Loftus, 1991).

Arntzen (1982, 1983) (cited in Steller & Köhnken, 1989), Undeutsch (1982, 1984), and Trankell (1972), on the other hand, argued that one cannot experimentally manipulate the veracity of statements with enough experimental realism in order to simulate real lies or truths. Specifically, they contended that one cannot experimentally mimic the emotional trauma of an actual child sexual abuse event, and therefore the typical experimental manipulation of a subject's statement as either truthful or deceptive cannot be used to generalize to real child sexual abuse allegations. They believed the CBCA method's validity, as well as its perfected use,

must come from the clinical practice of its application. Yuille (1988a) likewise criticized the use of simulated truthful or deceptive statements to test hypotheses in studies of children's eyewitness accuracy. He also argued that such studies are emotionally neutral and therefore flawed in that they do not equate to real traumatic events. Hence, these investigators felt that field studies were more appropriate than experimental studies to test the validity of the CBCA. For instance, Undeutsch (1982, 1984) endorsed the approach of comparing the "configuration of facts," in other words, the conclusions made by the criminal justice system, with the conclusions arrived at when applying the CBCA to the child's statement. If the two are congruent, the CBCA method is corroborated. Most of the field studies employ Undeutsch's (1982, 1984) approach. Besides experimental and field studies, there have been "collateral" studies that shed light on the CBCA. These studies, although not purposely focused on the CBCA technique, provide some support for its use.

Whether field, experimental, or collateral, many of the following studies are difficult to evaluate for two main reasons. Some are not published in refereed journals and only incompletely cited in secondary sources. This has resulted in a lack of authoritative details that would be useful in analyzing the methodological and statistical sufficiency of the studies. Additionally, some of the studies are published in English and some in German, but not both languages. This makes it difficult for non-bilingual researchers to get a complete assessment of the

CBCA's utility.

Field Studies

Investigators in Europe have use the CBCA to analyze sexual abuse cases for the courts. Littmann and Szewczyk (1983) and Szewczyk and Littmann (1982) compared the frequency of the 19 CBCA criteria in statements made by five- to 18-year-old witnesses in 123 sex offenses between 1965 and 1976. They classified the statements as truthful or deceptive using the CBCA criteria. Admission of lack of memory (criterion 15) was the only one of the CBCA criteria that was detected in at least half of the truthful statements. Yet, they found that many deceptive statements contained criteria. They also found that some of the CBCA criteria did not distinguish between truthful and deceptive statements. Unfortunately, the secondary source (Steller & Köhnken, 1989) that reported these results did not specify which criteria were not useful in distinguishing between truthful and deceptive statements.

Arntzen (1982, 1983) analyzed victim statements from 24,000 cases (92% of them being sexual abuse) of which 60-70% (about 15,600) were classified as truthful by Arntzen or his colleagues based on the CBCA. In more than 90% of these cases (about 14,000), the courts went along with the diagnoses of Arntzen. Of those, however, only 866 (6%) resulted in a confession by the perpetrator. The secondary source which reported Arntzen's efforts (Köhnken & Steller, 1988) failed to report what the court decisions were for the more than 13,000 other cases which were diagnosed as has having a truthful allegation.

Esplin, Boychuck, and Raskin (1988) (cited in Raskin & Esplin, 1991b) conducted the first field study of the CBCA. Although he did not elaborate, Steller (1989) claimed this was the first field study that met minimum scientific standards. Esplin et al. (1988) analyzed 40 child victim witness statements from sexual abuse cases. The children were between three and 15 years old. In all of these cases, the alleged perpetrator was acquainted with the child. Of the 40, 20 were considered confirmed based on a later confession by the perpetrator (18 cases) and/or by unequivocal evidence supporting the child's allegation (16 cases) (14 cases fulfilled both, two had only medical evidence, and four had only a confession). The other 20 were classified as doubtful based on either persistent denial by the accused, lack of corroborating evidence, recantation by the child, or judicial dismissal of the case. Also, polygraph results were available in 13 of the 20 doubtful cases and indicated the accused was truthfully denying the allegation. The two groups were similar with regard to gender, age, and type of sexual abuse. In the confirmed group, half of the alleged perpetrators were from the child's family. The doubtful group, however, reflected a different composition in this respect, with 18 of the alleged perpetrators coming from the child's family. Also, in the confirmed group, only two cases stemmed from a custody dispute, whereas in the doubtful group, 18 did so.

The statements (transcribed from audiotaped interviews) were then assessed by Boychuck, who had received training in the CBCA procedure and was blind to the group membership and facts

surrounding the cases. Boychuck applied each of the CBCA criteria to each of the statements and coded them with a "2" if a criterion was strongly fulfilled, with a "1" if a criterion was somewhat fulfilled, and with a "0" if a criterion was not fulfilled. When the points assessed for each criterion for each statement were added within the two groups, the transcripts that were classified as confirmed had a mean score of 24.8 points whereas the doubtful statements had a mean of 3.6 points. This was a statistically significant difference and there also was no overlap between the distributions of scores for the two groups. This demonstrated that truthful statements (as defined by the investigators) manifested more of the criteria. Table 4 shows the presence or absence of 19 CBCA criteria in the confirmed and doubtful statements. Seven of the criteria were present in every confirmed case (logical structure-criterion 1, quantity of details-criterion 3, contextual embedding-criterion 4, interactions-criterion 5, superfluous details-criterion 9, spontaneous corrections and additions-criterion 14, and characteristic details of the offense-previously criterion 19). Seven other criteria were absent in all the doubtful cases. These were criteria 6 (reproduction of speech), 7 (unexpected complications), 8 (unusual details), 11 (related external associations), 13 (attribution of accused's mental state), 16 (raising doubts about one's own testimony), and 17 (self-deprecation). But, the authors found that 55% of the doubtful statements contained logical structure and quantity of details. See also Table 7 for a summary of the criteria found to be useful

in this study.

Insert Table 4 About Here

In response to a criticism by Wells & Loftus (1991) that the doubtful group was defined too much by whether the authorities were convinced by the child's allegation, Raskin and Esplin (1991c) re-analyzed the above data of Esplin et al. (1988) after dropping two subjects who did not fit a new definition of doubtful allegations. This new definition was fulfilled if the case met two of the following: lack of medical evidence, recantation, or truthful polygraph of accused. This re-analysis showed the same evidence.

Boychuck (1991) applied the CBCA to the statements of 75 children (4-16 years old) who alleged sexual abuse during the 1987-1988 period. Fifteen of the children were boys and 60 were girls. Of these children, 64% were white and 9% were black. Transcripts were made from either audio or video taped police interviews. The statements were grouped into "confirmed" and "doubtful" in a similar fashion as in Esplin et al. (1988). Three raters evaluated the transcripts for the presence of CBCA criteria and found that criteria 1-8, 11, 12, 14, and 19 were present significantly more in confirmed than in doubtful transcripts. Boychuck also discovered that criteria [illegible] and 15 were present significantly more in older children than in younger children.

Anson, Golding, & Gully (1993) studied 23 cases of confirmed child sexual abuse in order to assess the CBCA's inter-rater

reliability. Anson et al. (1993) started out with 512 consecutive cases of children referred for sexual abuse assessments in Utah between 1986 and 1989. Cases were dropped from consideration if: authorities could not be contacted about the case; there were more than one alleged perpetrator; the allegation was made by someone other than the child victim; the accused did not confess; and the interview of the child was not videotaped. The classification of a case as confirmed if the alleged perpetrator later confessed is a much more conservative method than using a conviction or other judicial conclusions as indicative of a truthful allegation. The group of victims was comprised of 43% male and 57% female, four to 13 years old. Ninety-six percent of the children were white. Sixty-one percent of the alleged perpetrators were related to the victim.

The interviews were conducted by 10 different interviewers who did not use the same style of interview. Each videotape was rated by two of four trained judges for the presence of CBCA criteria. Table 5 shows reliability coefficients resulting from the four raters' ratings and the endorsement rates of the criteria. Three measures of inter-rater reliability were used: proportion agreement, Cohen's kappa, and Maxwell's Random Error (RE) coefficient. Cohen's kappa is a chance-corrected measure of proportion agreement, but it has been criticized because it is too responsive to the base rates of samples (Uebersax, 1987). Maxwell's RE coefficient was included in order to avoid overestimation of chance agreement that Cohen's kappa would produce

when the raters' base rates of presence or absence of a criterion are different from 50/50, as they are for some of the criteria (Maxwell, 1977). Notice in Table 5 that when the average presence of a criterion is near .50, Cohen's kappa and Maxwell's RE are similar.

Insert Table 5 About Here

The results showed that on average, the criteria were present only 41% of the time. This was in contrast to the Esplin et al. (1988) study which found a 70% occurrence rate. Among other reasons, the authors suggest that perhaps the fact that their study employed videotape instead of a statement transcript (which was the intended target of CBCA) or the low rate of free narrative interviews in their study may have led to the difference. More importantly, Anson and his colleagues found a wide range of reliability coefficients for the criteria depending on which criterion was analyzed. If one uses .75 as a minimally acceptable reliability coefficient, then only four criteria have adequate reliability. These were pardoning the perpetrator (criterion 18), doubting one's own testimony (criterion 16), perpetrator's mental state (criterion 13), and misunderstood details (criterion 10). Anson et al. (1993) classified the criteria into three categories based on the Maxwell's RE. These were: (1) Adequate reliability -- criteria 18, 16, 13, 10, 17, 6, 3, 1, 7; (2) Marginal reliability -- criteria 9, 4, 8, 14; and (3) Inadequate reliability -- criteria

15, 11, 5, 12, 2, and 19.

The authors identified some factors that may have affected the representativeness of the reliability coefficients. These were the inherent impreciseness of some criteria (e.g., interactions, details characteristic of the offense, and unstructured production), criteria based on a judgment of the child's behavior but lacking a standard (e.g., unusual details), infrequent occurring criteria (e.g., pardoning the accused, misunderstood details), and the ubiquitous nature of some criteria (e.g., logical structure). They also noted that the cases analyzed were all confirmed as true allegations of abuse. Unconfirmed or ambiguous cases may have even lower reliability coefficients. Lastly, the authors found a significant positive correlation between the age of the child (4-13 years old) and six of the criteria: logical structure (.52), contextual embedding (.63), descriptions of interactions (.66), reproduction of conversations (.50), pardoning the perpetrator (.43), quantity of details (.57), and the total CBCA reliability coefficient score (.54), indicating that older children exhibit the criteria more than younger children. However, there was no difference between the ratings of male children and female children. See Table 7 for a summary of the criteria Anson et al. (1993) found to be useful.

Experimental Studies

The first experimental study that evaluated the validity of certain CBCA criteria was by Köhnken and Wegener (1982). In their study, 37 female adolescents (16-17 years old) were recruited as

subjects. The teenage girls either viewed a 10-minute film of a family argument, or were told a brief synopsis of the event by the experimenters. The subjects were then interviewed (free narrative followed by structured questioning) about the event and were instructed either to relate it as they viewed it (film group), or construct a coherent story from what they were told (synopsis group). Transcripts of their interviews were then coded by five blind judges for unstructured production (criterion 2) and the amount of detail (criterion 3), plus "consistency of the content over repeated questioning," which is another part of the SVA. The experimenters hypothesized that the statements from the girls who observed the filmed event would have more details and more unstructured production. The results supported the hypothesis with regard to criterion 3 but not criterion 2: falsified stories exhibited unstructured production significantly more than truthful stories.

Yuille (1988b) (cited in Steller, 1989) had 49 first- and third-graders tell both a true and false story about something that happened to them. After two days of rehearsing the stories, they were interviewed according to SVA guidelines by a judge who was blind to the study. After that, transcripts of their stories were independently coded by an additional two judges for the presence of the CBCA criteria. The children did not always do as were instructed, i.e., tell one true and one false story. But, when looking at all the true stories actually told and all the false stories actually told, these judges accurately identified 91% of

the true stories and 74% of the falsified ones. Other than supporting the utility of the CBCA, this study suggested that the CBCA was more susceptible to classifying false statements as true than classifying true statements as false. The inter-rater reliability in this study was very good, as two judges who classified the statements as truthful or false were in agreement 96% of the time.

Steller, Wellershaus, and Wolf (1988) (cited in Steller, 1989) found further support for the CBCA criteria in distinguishing between true and invented stories of 98 first- and fourth-graders. These authors attempted to better simulate the emotional and experiential context in which real-life sexual abuse allegation are made by picking experimental scenarios in which 1) the children were directly involved in, 2) the child was negatively emotionally aroused, and 3) the event in question was characterized by a loss of control by the child. In this study, the children were asked to relate both a true and a false story. The stories were to consist of either receiving an injection, undergoing an operation, having a blood sample taken, having dental work performed, suffering an accident requiring medical treatment, being beaten up by another child, or being attacked by a dog or animal.

To assess the reliability of the criteria, three undergraduate psychology students were given a 90-minute training session on rating statements for the CBCA criteria on a four-point scale. This scale ranged from 0 (criterion not present) to 3 (criterion strongly present). In assessing 194 of these stories (two were not

rated because they were used during the training session), these raters disagreed very little (disagreement was defined as one or two of the raters making a "not present" choice on one criterion and the other rater making any other choice). The modal disagreement among raters was 2%, with the highest disagreement being 7% for criterion 7 (unexpected complications). There was no disagreement on criteria 1, 2, and 3. The three undergraduate students' ratings of the stories were subjected to a series of ANOVAs to see if any criterion was rated significantly higher in true stories than in false ones. Not counting criteria 19 (offense-specific) and 16 (raising doubts about one's own testimony) which were eliminated because they either were not relevant to the stories (criterion 19) or none of the raters assessed it (criterion 16), the three trained students detected 11 of the CBCA criteria significantly more in truthful statements than in falsified statements. Criteria 2, 13, and 14 through 18 did not significantly distinguish between true and false stories. These ANOVA results applied only for the stories that were medical in nature (n = 127). As reported by Steller (1989), a "first analysis revealed that the Undeutsch hypothesis could only be corroborated for stories with medical topics. . . ." Steller et al. (1988) speculated that the CBCA criteria would distinguish between true and false accounts only when the subject matter was "more or less intimate manipulation of the body of the person affected." Unstructured production, accused's mental state, and all the motivation-related criteria failed to distinguish the false from

true stories.

In order to determine the usefulness of the CBCA in helping to classify truthful and falsified stories, the authors then compared the classifications of twenty-five undergraduate psychology students who had no knowledge of the CBCA and the three undergraduates who received the 90-minute training session. These two groups of students rated a sub-sample of about 40 of the 196 stories on a five-point scale from 1 = very unlikely untrue to 5 = very likely true. The results showed that CBCA training enabled correct classification of truthful (78% of the time) and false (62% of the time). The 25 untrained raters were able to correctly classify truthful stories 68% of the time and false stories 47% of the time. These differences were statistically significant using chi square tests. The ability of the untrained subjects to correctly classify 68% of the true stories suggests some factors other than CBCA training may have had an effect in interaction with the training to aid in classifying the truthful stories.

Like the results in the Yuille (1988b) study, this study found that errors associated with the use of the CBCA tend to be false positives rather than false negatives. In other words, there was a higher tendency to falsely classify fictitious statements as truthful than to falsely classify true statements as fabricated. Steller (1989) considered this outcome as commensurate with the development and purpose of the CBCA. That is, it was specifically developed in Germany at a time when children's testimony was generally regarded as incredible, so its purpose was to corroborate

credibility, not to detect deception.

In another experimental study, Landry and Brigham (1992) trained 114 college students in the CBCA technique (only 14 of the 19 criteria were used; 3 because they specifically related to children and 2 because pilot work demonstrated that subjects were unable to apply the criteria consistently). They also enlisted the aid of 70 undergraduates to act as senders. The senders were videotaped giving either a true or fabricated 1- to 2-minute description of personal incidents that were traumatic, emotional, and involved feelings of a loss of control (these are two of the three factors suggested by Steller et al. (1988) to more closely simulate sexual abuse trauma). Seven raters evaluated the 140 videotaped stories for their emotionality, trauma, and loss of control. Twelve of the stories were selected based on their moderate ratings given by these seven raters. The topics of the true stories were death of family member (2 stories), cancer in the family, near victim of date-rape, victim of a burglary, and family member committing suicide. The fabricated stories consisted of family member died in auto accident, pregnancy, victim of burglary, abortion (2 stories), and victim of a date-rape.

All 12 stories were presented to all subjects. They were presented to either subjects trained in the CBCA (45-minute session) or untrained subjects. The stories were presented either in the form of videotape or only the written transcript of the videotape. They found that only when CBCA training was coupled with viewing the videotape, were the subjects able to correctly

classify the stories as either truthful or fabricated significantly better than chance, and then, not much better than chance. When trained and viewing the videotape, on average, the subjects accurately classified the stories 58% of the time. The effect of training appeared to be on the classification of true senders. Untrained subjects were able to correctly identify true stories only 59% of the time, whereas trained subjects did so 75% of the time. But, both trained and untrained subjects had a poor ability to correctly classify false stories (35%). This finding is unsettling and again shows that the CBCA has a tendency to make false positive classifications more than it does false negatives.

Further analyses showed that 10 of the 14 criteria were present significantly more in true stories than in false ones (See Table 6). Unexpectedly, criteria 1 (logical structure) and 13 (attribution of accused's mental state) were seen more frequently in the false stories than the true ones. Additionally, criteria 7 (unexpected complications) and 17 (self-deprecation) did not distinguish between true and invented stories.

Insert Table 6 About Here

Like the Steller et al. (1988) study, this study also suggests that non-verbal factors may have had an effect in interaction with CBCA training. Unfortunately, these authors did not separate transcript-only from videotape stories when reporting the extent to which criteria discriminated between truthful and false stories.

Also, this study demonstrated that there was no effect of the speaker's gender on the ability to classify them accurately.

Joffe and Yuille (1992) (cited in Horowitz, 1991; Köhnken, Schimossek, Aschermann, & Hofer, 1993) demonstrated that using CBCA criteria-rich material to coach a deceiver about an event may have made it difficult for CBCA trained judges to later distinguish between fabricated transcripts and truthful transcripts. This suggests that if sexual abuse allegations are the result of extensive prodding and coaching by adults, the CBCA method may not be as useful.

Hofer, Köhnken, Hanewinkel, and Bruhn (1992) (cited in Köhnken et al., 1993) showed a film of a robbery to 56 adult subjects and then asked the subjects to describe the event truthfully or in a predetermined distorted manner. A subset of CBCA criteria were successfully used to distinguish truthful from deceptive accounts. A discriminant analysis in this study showed that the factor "description of unexpected complications" had the greatest value in discriminating between the two groups.

Köhnken et al. (1993) found additional support for the CBCA utility even when a cognitive interview instead of a free recall interview style was used. Fifty-nine adult interviewees either saw a film of a person giving blood or were told about the film. They were then interviewed by one of six interviewers trained on part of the CBCA. Criteria 4, 11, 12, 17, and 19 were not used. Also, criteria 5 and 6 were combined into a criterion called "description of verbal and non-verbal interactions." The film group gave a

truthful account and the other group was instructed to describe to the interviewer the events of the film as if they actually saw it. Their interviews were later transcribed and the CBCA criteria were rated as present or not. A MANOVA with truthfulness of statement as an independent variable and the interviewers' ratings of the presence or absence of CBCA criteria as the dependent variables showed a significant effect for the independent variable, indicating the presence of the criteria in the truthful statements was higher than in the false statements. But, only criterion 3 (quantity of details) and criterion 2 (unstructured production) were found in a univariate analysis to be significantly more present in true statements than in false statements. A discriminant analysis also demonstrated the ability of the CBCA criteria to distinguish between the two groups. A significant discriminant function was found for the criteria and the classification accuracy was 85%. Köhnken et al. (1993) found that additional criteria, such as insecurities, cliches, and reporting style, contributed to distinguishing between true and false statements, however, they felt it was premature to include them in the CBCA at that time.

Assessing the Empirical Foundation of the CBCA

The attempts to assess the validity of the CBCA as outlined above have had mixed results. No one study can possibly take into account all possible variables that may influence the usefulness of the CBCA. Taken as a whole though, the few studies that have been conducted provide convergent validity for the CBCA method and it

appears that the CBCA can be a useful tool to aid in determining if an allegation is truthful or falsified. For example, Tables 7 and 8 provide summaries of the CBCA criteria which have been shown to have some usefulness in distinguishing truthful from false statements. There are some important considerations, however, that must be addressed in assessing past research and in conducting future studies.

Insert Tables 7 & 8 About Here

Ambiguity of Studies

As mentioned earlier, one of the most serious impediments to an analysis of the CBCA's utility, is the manner in which some studies were reported. Unfortunately, most of the earlier studies on the CBCA are in German and have not been translated. What is known of these studies (at least for non-bilingual researchers) is limited to secondary sources. Likewise, some of the studies that are in English have either not been published in refereed journals or are not easily accessible. Just as with the German studies, this makes it necessary to rely on secondary sources for analysis. A sufficient assessment of the studies' methodology is difficult since details are lacking.

Strength of Empirical Support

Littmann and Szewczyk (1983) and Szewczyk and Littmann (1982) only demonstrated weak evidence for the validity of CBCA, namely, that criterion 15 was present in at least half of the truthful

statements. These authors also showed that many of the criteria did not distinguish between truthful and deceptive statements. Köhnken & Wegener (1982) likewise found only minimal support for the discriminating ability of the CBCA criteria. In their study, only criterion 3 did so. With a little more support, Köhnken et al. (1993) found that criteria 2, 3, and 6 were present in truthful statements significantly more than in false ones. On the other hand, four of the studies showed substantial support for the utility of the criteria (Anson et al., 1993; Boychuck et al., 1989; Steller et al., 1988; and Landry & Brigham, 1992).

In addition to demonstrating a higher presence of some criteria in truthful statements than deceptive ones, three studies provided substantial evidence for the ability of the CBCA method to assist in accurately classifying truthful and false statements (See Table 8). Yuille (1988b) found that CBCA training enabled judges to accurately identify 91% of true stories and 74% of falsified ones. Steller et al. (1988) found less accuracy rates, but still showed the usefulness of the CBCA training in detecting truthful statements 78% of the time and false ones 62% of the time. In that study, though, even the untrained raters were able to correctly classify truthful stories 68% of the time, suggesting some factors other than CBCA training were involved. Landry & Brigham (1992) obtained results that showed no effect of CBCA training on detecting false statements. They achieved a 75% accuracy rate when classifying true statements but only 35% when classifying false statements. The Landry and Brigham (1992) figures, however,

include both raters who used transcripts alone and raters who also viewed videotapes of the speakers. Similar to the finding in Steller et al. (1988), untrained subjects in the Landry and Brigham (1992) study were able to correctly identify true statements 59% of the time.

Reliability of the Presence of Criteria

A few studies that addressed inter-rater reliability found high rater agreement when analyzing the overall ratings of truthfulness of children's statements and when analyzing individual criteria. But, some criteria were also found to have marginal or inadequate reliability. Yuille (1988b) obtained a .96 inter-rater reliability on overall ratings when two judges used CBCA criteria in rating 49 children's statements as true or false. With respect to specific criteria reliability, Anson et al. (1993) demonstrated that only nine of the CBCA criteria were adequately reliable (as defined by the authors as above .56); 10 of the criteria were either marginally or inadequately reliable. In Steller et al. (1988) three judges obtained higher inter-rater reliability estimates of between .93 and 1.00 for each criterion when rating 194 children's statements. There are still too few studies which addressed the question of reliability of the CBCA as a diagnostic tool.

Selection Bias

As noted by Steller and Köhnken (1989), field studies thus far have been highly selective in the cases they chose to analyze, using external adjudicative factors such as confession or

conviction to assume a statement as truthful. For instance, in Esplin et al. (1988) and Boychuck (1991), the final court decision was one of the standards used to classify a statement (e.g., if the judge dismissed the charges, the allegation was classified as "doubtful"). But, a dismissal might be more a result of the child's ability to convince and/or the system's willingness to undergo a trial, not necessarily the truthfulness of the child's statement. Polygraph results and confessions are probably more indicative of the truthfulness of a child's allegation. In Anson al. (1993), similar selection bias occurred. The age of the child can also compound this selection bias to the extent that younger children may not have the communicative or persuasive skills as do older children and, therefore, younger children's allegations may be at risk of being believed less than those of older children. There in fact might be a positive correlation between some of the CBCA criteria and a child's age, with older children manifesting more of the certain criteria (Anson et al., 1993). This may be due to older children being able to display these criteria more than young children when interviewed. The selection process used by Anson et al. (1993) and Esplin et al. (1988) might have resulted in one group consisting of convincing, articulate, and older children (rather than as truthful), and the other group consisting of unconvincing, less articulate, and younger children (rather than as deceivers).

Another possible selection problem may result in a circular process. If professionals (judge, jurors, law enforcement)

unwittingly use some of the CBCA criteria when judging truthfulness of a statement, as was found in Coolbear (1991), then resulting adjudicative conclusions would be based on the absence or presence *of some* criteria. Subsequently applying the CBCA method to the statement would therefore necessarily result in the detection of criteria and the possible classification in line with the CBCA.

Further, even if these external factors accurately distinguished between truthful and falsified statements, they probably identified very clear and unambiguous ones (Wells & Loftus, 1991). Any results may not generalize to ambiguous cases. In order to increase the ability to select cases that can clearly fit into "confirmed" and "doubtful" categories, investigators typically use only cases that are not ambiguous. Anson et al. (1993), for example, started with 512 cases and weeded out those that did not fit certain criteria, in other words, those which were ambiguous. Therefore, they are using only those cases that are on the extreme ends of a continuum with strong indication of truthfulness on one side and strong indication of falsehood on the other. The ambiguous cases (the ones that would benefit the most from the CBCA) are not used. For practical reasons, some selection biases may be an unavoidable weakness of field studies. Nonetheless, increased attention to this potential weakness and added emphasis on experimental designs can address the problem.

Transcript v. Statements v. Audio/Video

In Anson et al. (1993), Hofer et al. (1992), and Steller et al. (1988), it is unclear whether written transcripts or videotapes

of interviews were analyzed. In Landry and Brigham (1992) the subjects rated videotapes in addition to transcripts. In fact, Landry and Brigham (1992) found that only when viewing videotape, CBCA trained raters were able to accurately classify statements beyond the chance level.

The CBCA was intended to analyze the verbal content of an allegation. Although the verbal content can of course be gleaned by viewing a video or audiotape, there most likely are other non-verbal factors (visual or auditory) that the perceiver would also either intentionally or unwittingly use to make a judgement of veracity. Any apparent effect of CBCA training would then be contaminated by the effect of these other non-verbal variables.

Also, a few of the studies (Arntzen, 1982, 1983; Littmann & Szewczyk, 1983; Szewczyk & Littmann, 1982) were unclear whether formal written statements from the witnesses were analyzed rather than written transcripts of their interviews. The process of police or investigators taking a formal statement from a witness could lend itself to the incorporation of the interviewer's choice of words in order to fulfill anticipated prosecutorial criteria. Therefore, the process might result in a statement that does not reflect the true verbal content of the person making the allegation. This would make the statement an inappropriate indicator of the witness' veracity. Future research should be careful to employ verbatim transcripts of a speaker's statement rather than video and/or audiotape or formal statements.

Control Groups

Of the three studies which reported the effect of CBCA training on a subject's ability to correctly classify true and false statements (See Table 8), only Landry and Brigham (1992) and Steller et al. (1988) used untrained control groups. Without control groups, it is impossible to determine if a rater's ability to distinguish between true and false statements is an effect of the CBCA training or the non-verbal cues from audio or video presentation. Also, having access only to the verbal content of a speaker's message in the form of a written transcript might enhance one's ability to detect deception in the statement (see Köhnken, 1989b). In order to control for this possible contaminating effect, studies must employ control groups that account for verbal and non-verbal cues to deception and truthfulness.

Lies v. Confabulation or Fantasy

The research on the CBCA has not distinguished between false stories in which children are deliberately lying and false stories in which children are relating information which was suggested to them or which they have fantasized, but which they believe to be true. The question is, will the CBCA be sensitive to false information the child believes is true (information the child assimilates through suggestibility or information the child fantasizes) as it is to information the child knows is a lie. The main hypothesis of the CBCA states that a truthful statement will differ in quality from a falsified one due to a truth-teller relying on the real memory of the event. If this is accurate, then

as long as the child has either a real memory of the event, or a pseudo-memory created by fantasy or suggestion, the CBCA method should classify the child as truthful. In other words, the CBCA should not be sensitive to these pseudo-memories as falsehoods. This issue should be addressed in future research.

Age and Developmental Issues

The CBCA was specifically developed with the child witness in mind. The 1954 ruling by the Supreme Court of the Federal Republic of Germany mandated expert assessment of the veracity of child sexual abuse victims (Raskin & Yuille, 1989), and was probably a significant catalyst in subsequent development of the CBCA. However, some experimental studies have been conducted using adults as speakers. Raskin and Esplin (1991a) expressed caution in using the CBCA with adolescents and older individuals. Their concern mainly rested with adolescents' increased knowledge of sexual matters, which would presumably enable them to give more detailed description even if lying, and increased manipulative and control skills.

Having said that, it may not be important to use children in testing the validity of the CBCA. The criteria apparently work with adult speakers. Hofer et al. (1993), Köhnken et al. (1993), and Landry and Brigham (1992), found usefulness of the CBCA with adults. Also, as shown in Landry and Brigham (1992) and Köhnken et al. (1993), it is the relative presence of CBCA criteria between truthful and false statements that distinguish them, not their absolute presence or absence. To find an adult truthful, one may

merely need a higher threshold score than for children. The CBCA may be a useful tool in discriminating between truthful and deceptive statements of any witness.

Additionally, there may be developmental concerns when looking only at children. Piaget's developmental stage theory provides some support for this possibility. Changes between the preoperational, concrete operational, and formal operational stages (those during the ages when children are likely to make sexual abuse allegations), suggest qualitative differences in a child's verbal content (Siegler, 1991). Preoperational children (2-7 years old) can only view the world from their own perspective, focusing their attention too narrowly and often ignoring important information. Although children in the concrete operational stage (7-12 years old) can understand other people's perspectives, they still do not consider all possible outcomes of actions. Formal operational children (above 12 years old) have a grasp of abstract concepts. These developmental changes in verbal abilities are likely to affect verbal content and result in either more or less of certain criteria. In fact, Boychuck (1991) found that the statements of older children apparently contained more content associated with criteria 5 and 15. Further research should address this developmental issue.

Nature of Message

In a similar fashion to the previous issue, the CBCA was developed specifically to be applied to sexual abuse allegations. Just as studies have not exclusively used child speakers, some

experiments have not employed sexually explicit themes in their speakers' statements. For obvious ethical considerations children could not be expected to relate actual or contrived incidents of being sexually abused. Even with adults, it would be difficult to convince an ethics review board to allow them to relate authentic incidents of sexual abuse. Steller et al. (1988) suggested that it is sufficient for the information to be related to be negatively emotional, involve a loss of control, and directly involve the speaker. Subsequent experiments have demonstrated that the CBCA technique is useful in determining whether statements containing information of this nature are true or false.

Yet, if the CBCA criteria are considered to be indicative of the effect of deception on verbal content, the content's theme may not matter. The results of some of the previously outlined studies suggests this is so and that the CBCA can aid in detecting true statements that are not related to child sexual abuse or even related to the three criteria proposed by Steller et al. (1988). Recall that Undeutsch's (1982) hypothesis was that truthful statements are qualitatively different from invented ones in that truthful ones were actually experienced by the person. This hypothesis does not consider the nature of the message. Also, the additional emotional or conflicting nature of deception (i.e., the affective difficulty stemming from relating what one knows not to be the truth), especially if motivated to be successful in deceiving, should be sufficient in producing the alteration in verbal content. This issue should be considered when planning

future studies.

Motivation of Speakers

As pointed out by DePaulo, Lanier, and Davis (1983), raters have a more difficult time detecting deception when the speaker is motivated to avoid detection. This is only true, however, when the raters are limited to evaluating the speakers' verbal content. The field studies above obviously used actual statement content of the alleged victim, so the motivation of the speakers was authentic. However, in the experimental studies, speakers knew they were presenting deceptive or truthful statements for experimental purposes and were unlikely to be experiencing much motivation to deceive the subjects. One exception might be that child speakers were told that the experiment was part of a game and may have felt some incentive to convince the interviewer of their story. But, they were definitely not provided with any strong incentive (such as monetary or social reward) to deceive. As earlier stated, Arntzen (1983), Horowitz (1991), Trankell (1972), Yuille (1988a), and Undeutsch (1982, 1984) expressed similar reservation about the ability to have a sufficiently motivated deceiver when conducting experimental studies. Although it may be difficult, researchers should plan for motivating their speakers in the future either with monetary rewards for successful deception or other social incentives.

Cultural Differences

Perhaps the most serious criticism of previous studies on the CBCA concerns cultural differences. The application of any method

to detect truthfulness and deception based on a person's verbal content (as the CBCA does) may be influenced by such differences. The existence of cultural and/or racial differences between the European context (in which the CBCA was developed) and the American context suggests the need to assess the effect of such differences. The studies conducted on the CBCA have primarily been with European and/or white individuals making statements. For instance, in Boychuck (1991), only 9% of the statements were from black children. In Anson et al. (1993), 96% of the sexual abuse cases studied were of white children. In Landry and Brigham (1992), all subjects and senders were white. None of the other authors reported the race of the children in their studies. In telephone consultations in 1992 with the U.S. Bureau of Justice Center for Child Abuse and Neglect, the National Committee for the Prevention of Child Abuse, and the Florida Child Abuse Registry, personnel at these agencies indicated that cross-race evaluations occur frequently where the child victim of sexual abuse is interviewed by someone of another race (Brigham, personal communication). Yet none of the studies addressing the utility of the CBCA have looked at its application to different race speakers or when applied by different race subjects. Hall (1959) said that non-verbal behavior is culturally learned and must be interpreted with knowledge of the other culture in order for one to be accurate in interpreting it. With respect to verbal behavior, Murray (1983) conducted a study which suggested that judges need prior familiarization with a speaker's repertoire during truthful communication in order to

accurately recognize a deceptive message. In further support of a same-race bias in detecting deception, Nance (1993) found that black and white observers were most successful in detecting deception when the target person was of the same race.

The fact that no research has looked at this is especially disturbing since not only may cultural sensitivity affect the ability to detect deception in cross-race situations, but also race-related schemas and biases of subjects may interfere with an analysis of the statement of a speaker. Surveys indicate that 10% to 15% of reported sexual assault cases involve a child and perpetrator of different races (Geiser, 1979). These cross-race child sexual abuse cases might provoke more prejudicial reactions by those who have strong race-related schemas. Although there is not much research on the cross-race evaluations of children, there is ample evidence that racial schemas can affect criminal justice decisions (Johnson, 1985; Pfeifer, 1990; Ruby and Brigham, 1993) and, in general, be used in interpreting information about members of a certain racial group (Dovidio, Evans, & Tyler, 1986; Higgins & Bargh, 1987; Wyer, 1988).

Conclusion

Within the last decade, detecting child sexual abuse has become a priority among politicians, legislators, law enforcement officials, and laypersons. Numerous allegations have been made which rest solely on the victim's testimony; it is many times the only evidence that the authorities have on which to base prosecution or intervention. In an attempt to protect the

interests of both the child and the alleged perpetrator, the CBCA was developed to address the victim's veracity based on the content of his or her verbal allegation.

For most of the legal system's history, it has distrusted the veracity of these allegations. However, in recent years, many institutions involved in responding to sexual abuse allegations shifted their stance and are more inclined to consider the allegations true (Coolbear, 1991; Faller, 1984). The courts, however, are still doubtful as to a abuse victim's truthfulness and its potential impact on the defendant's freedom. As a result, the courts are careful in admitting a victim's testimony, and have been specific in their preference for scientifically supported methods of assessing the child's veracity (See Rozell, 1985; State v. Kim, 1982).

The CBCA has been heralded as such as method. Although the research is sometimes only accessible through secondary sources, and many variables have yet to be addressed, researchers have successfully provided empirical data to bolster the utility of the CBCA. Those studies conducted have either not been experimental in nature and fall prey to selection bias, or have fallen short of addressing several important issues.

First, it appears that investigators have not been careful to apply the CBCA to written transcripts of the verbal content of the alleged victims. They have sometimes allowed non-verbal behavior to get through to the raters via video/audiotape. Second, control groups have not been used enough when using the CBCA to classify

true and false statements. This weakness has prevented an adequate analysis of whether CBCA training is effective, or whether non-verbal and verbal cues from video/audio tapes or written transcripts are the important factor. A third consideration is the failure of studies to differentiate between motivated lies, suggested falsehoods, and fantasies of the child. It is not yet clear whether the CBCA would consider all three as deceptive. Since the child actually believes the latter two, would the CBCA classify them as truthful? Fourth, experiments using the CBCA on adults have been successful in demonstrating its utility. This suggests that the use of the technique may be expanded. The fifth and sixth issues relate to the nature of the message being spoken and the motivation of the speaker. The CBCA may be appropriate for non-sexual abuse matters as long as the emotional, personal, and lack of control characteristics are maintained. Also, motivated speakers do better at deceiving, so research subjects should be motivated as much as possible to better simulate real deceivers, who are obviously motivated. Lastly, cultural differences between the speaker and rater need to be fully addressed in future research. Past CBCA studies have without exception ignored the different types of verbal content that may be manifested by different cultural and racial groups, and the possible different abilities to accurately detect truth-tellers and deceivers in cross-cultural situations.

Future research on the CBCA must address these issues so that the CBCA can fit within the parameters of allowed judicial

procedures and provide the criminal justice community with a tool to use in making decisions about allegations of child sexual abuse and possibly other crimes. The CBCA has the potential to enhance the objectivity of the investigation and prosecution of allegations of child sexual abuse and it can also aid in protecting those unfortunate enough to be at the receiving end of a child sexual abuse allegation.

REFERENCES

American Law Reports Index. (1992). New York: Lawyers Cooperative Publishing.

Anson, D. A., Golding, S. L., & Gully, K. J. (1993). Child sexual abuse allegations: Reliability of criteria-based content analysis. Law and Human Behavior, 17, 331-341.

Arntzen, F. (1970). Psychologie der zeugenaussage. Einfuehrung in die forensische aussagepsychologie. Goettingen: Hogrefe.

Arntzen, F. (1982). Die situation der forensischen Aussagepsychologie in der bundesrepublic Deutschland. In A. Trankell (Ed.), Reconstructing the past: The role of psychologists in criminal trials (pp. 107-120). Stockholm: Norstedt.

Arntzen, F. (1983). Psychologie der zeugenaussage. Systematik der glaubwuerdyhietsmerhmale. Muenchen: C. H. Beck.

Baker, E. V. (May, 1990). Psychological expert testimony on a child's veracity in child sexual abuse prosecutions. Louisiana Law Review, 50, 1039-1055.

Bienen, L. B. (Spring, 1983). A question of credibility: John Henry Wigmore's use of scientific authority in Section 924a of the Treatise on Evidence. California Western Law Review, 19, 235-268.

Binet, A. (1900). La suggestibilite. Paris: Schleicher Freres.

Boychuck, T. D. (1991). Criteria-Based Content Analysis of children's statements about sexual abuse: A field-based validation study. Doctoral Dissertation. Arizona State University.

Bulkley, J. A. (1989). The impact of new child witness research in sexual abuse prosecutions. In S. J. Ceci, D. F. Ross, M. P. Toglia (Eds.), Perspectives on children's testimony (pp. 208-229). New York: Springer-Verlag.

Ceci, S. J., & Bruck, M. (1993). The suggestibility of the child witness: A historical review and synthesis. Psychological Bulletin, 113, 403-439.

Coleman, L, & Clancy, P. E. (Fall, 1990). False allegations of child sexual abuse. Criminal Justice, 5, 14 (12).

Commonwealth v. Seese, 517 A. 2d 920 (Pennsylvania, 1986)

Coolbear, J. L. (1991). Credibility of young children in sexual abuse cases: Assessment strategies of legal and human service professionals. Canadian Psychology, 33, 151-164.

Corder B. F., & Whiteside, R. (1988). A survey of jurors' perception of issues related to child sexual abuse. American Journal of Forensic Psychology, 6, 37-42.

DePaulo, B. M., & Pfeifer, R. L. (1986). On the job experiences and skill in detecting deception. Journal of Applied Social Psychology, 16, 244-267.

DePaulo, B. M., Lanier, K., & Davis, T. (1983). Detecting the deceit of the motivated liar. Journal of Personality and Social Psychology, 45, 1096-1103.

DePaulo, B. M., Stone, J. L., & Lassiter, G. D. (1985). Deceiving and detecting deceit. In B. R. Schlenker (Ed.), The self and social life (pp. 323-370). New York: McGraw-Hill.

DePaulo, B. M., Zuckerman, M., & Rosenthal, R. (1980). Detecting deception: Modality effects. In Wheeler, L. (Ed.), Review of Personality and Social Psychology, Vol I, Beverly Hills, CA: Sage.

Dettenborn, H., Froehlich, H., & Szewczyk, H. (1984). Forensische Psychologie. Berlin: Deutscher Verlag der Wissenschaften.

Dovidio, J. F., Evans, N., & Tyler, R. b. (1986). Racial stereotypes: The contents of their cognitive representations. Journal of Experimental social Psychology, 22, 22-57.

Ekman, P., & O'Sullivan, M. (1991). Who can catch a liar? American Psychologist, 46, 913-920.

Esplin, P. W., Boychuck, T., & Raskin, D. C. (1988, June). A field validity study of criteria-based content analysis of children's statements in sexual abuse cases. Paper presented at the NATO Advanced Study Institute on Credibility Assessment in Maratea, Italy.

Faller, K. C. (1984). Is the child victim of sexual abuse telling the truth? Child Abuse and Neglect, 8, 473-481.

Faller, K. C. (1988). Criteria for judging the credibility of children's statements about their sexual abuse. Child Welfare, 67, 389-395.

Geiser, R. L. (1979). Hidden victims: The sexual abuse of children. Boston: Beacon Press.

Green, A. H. (1986). True and false allegations of sexual abuse in child custody disputes. Journal of the American Academy of Child Psychiatry, 25, 449-456.

Hall, E. T. (1959). The silent language. Garden City, NY: Doubleday & Co.

Hawkins v. State, 326 So. 2d 229 (Florida, 1976).

Higgins, E. T., & Bargh, J. A. (1987). Social cognition and social perception. Annual Review of Psychology, 38, 369-425.

Hofer, E., Köhnken, G., Hanewinkel, R., & Bruhn, C. (1993). Dianostik und Attribution von Glaubwurdigkeit. Final report for research grant No. Ko 882/4-1 of the Deutsche Forshungsgemeinschaft. Kiel: University of Kiel.

Horowitz, S. W. (1991). Empirical support for Statement Validity Assessment. Behavioral Assessment, 13, 293-313.

Joffe. R., & Yuille, J. C. (1992). Criteria-based content analysis: An experimental investigation. Paper presented at the Meeting of the American Psychology-Law Society, San Diego.

Johnson, S. L. (1985). Black innocence and the white jury. Michigan Law Review, 83, 1611-1708.

Köhnken, G. (1982). Sprechverhalten und glaubwuerdigkeit: Eine experimentelle studie zur extralinguistischen und textstilistischen anssageanalyse. Unpublished doctoral dissertation, University of Kiel.

Köhnken, G. (1987). Training police officers to detect deception: Does it work? Social Behavior, 2, 1-17.

Köhnken, G. (1989a). The evaluation of statement credibility: social judgement and expert diagnostic approaches. Paper presented on the International conference on Children's Evidence in Legal Proceedings, Selwyn College, Cambridge.

Köhnken, G. (1989b). Behavioral correlates of statement credibility: Theories, paradigms, and results. In H. Wegener, F. Losel, & H. Jochen (Eds.), Criminal behavior and the justice system: Psychological perspectives (pp. 271-289). New York: Springer.

Köhnken, G., & Steller, M. (1988). The evaluation of the credibility of child witness statements in the German procedural system. Issues in Criminological and Legal Psychology, 13, 37-45.

Köhnken, G., & Wegener, H. (1982). Zur glaubwuerdigkeit von zengenaussagen. Experimentelle ueberpruefung ausgewaehlter glaubwuerdigkeitskriterien. Zeitschrift fuer Experimentelle und Angewandte Psychologie, 29, 92-111.

Köhnken, G., Schimossek, E., Aschermann, E., & Hofer, E. (1993). The cognitive interview and the assessment of the credibility of adult statements. Unpublished manuscript.

Kraut, R. E., & Poe, D. (1980). On the line: The deception judgements of customs inspectors and laymen. Journal of Personality and Social Psychology, 39, 784-798.

Landry, K. L., & Brigham, J. C. (1992). The effect of training in criteria-based content analysis on the ability to detect deception in adults. Law and Human Behavior, 16, 663-675.

Littmann, E., & Szewczyk, H. (1983). Zu einigen kriterien und ergebuissen forensisch-psychologischer glaubwuerdykertsbegutachtang von sexuall missbrauchten kinden und jugendlichen. Forensia, 4, 55-72.

Masson, J. M. (1984). The assault on truth: Freud's suppression of the seduction theory. New York: Farrar, Straus, & Giroux.

Maxwell, A. E. (1977). Coefficients of agreement between observers and their interpretation. British Journal of Psychiatry, 130, 79-83.

Murray, J. J. (1983). The detection of interviewers; verbal deception from their accompanying overt non-verbal behavior. The Florida State University.

Nance, C. (1992). The effect of specificity of experience on the ability to detect deception in same- and other-race target persons. Unpublished master's thesis. Florida State University.

Nicholas, H. A. (Summer, 1988). Criminal law - child sexual abuse - expert testimony as to credibility. Duquesne Law Review, 26, 997-1010.

Nurcombe, B. (1986). The child witness: Competency and credibility. Journal of the American Academy of Child Psychiatry, 25, 473-480.

Pfeifer, J. E. (1990). Reviewing the empirical evidence on jury racism: Findings of discrimination or discriminatory findings? Nebraska Law Review, 69, 230-250.

Quinn, K. M. (Spring, 1988). The credibility of children's allegations of sexual abuse. Behavioral Sciences and the Law, 6, 181-199.

Raskin, D. C., & Esplin, P. W. (1991a). Statement validity assessment: Interview procedures and content analysis of children's statements of sexual abuse. Behavioral Assessment, 13, 265-291.

Raskin, D. C., & Esplin, P. W. (1991b). Assessment of children's statements of sexual abuse. In J. Doris (Ed.), The suggestibility of children's recollections (pp. 153-164). Washington, DC: American Psychological Association.

Raskin, D. C., & Esplin, P. W. (1991c). Commentary: Response to Wells, Loftus, and McGough. In J. Doris (Ed.), The suggestibility of children's recollections (pp. 172-176). Washington, DC: American Psychological Association.

Raskin, D. C., & Yuille, J. C. (1989). Problems in evaluating interviews of children in sexual abuse cases. In M. P. Toglia, S. J. Ceci, & D. F. Ross (Eds.), Adult perceptions of child testimony (pp. 184-207)

Rozell, S. (Winter, 1985). Are children competent witnesses?: A psychological perspective. Washington University Law Quarterly, 63, 815-829.

Ruby, C. L., & Brigham, J. C. (1992). Survey of law enforcement officers' opinions about deception. Unpublished survey. Florida State University.

Ruby, C. L., & Brigham, J. C. (1993). A criminal schema: The role of chronicity, race, and socioeconomic status on law enforcement officials' perceptions of others. Master's thesis. Florida State University.

Schooler, J. W., Gerhard, D., & Loftus, E. F. (1986). Qualities of the unreal. Journal of Experimental Psychology: Learning, Memory, and Cognition, 12, 171-181.

Scott, W. T. (Spring, 1989). The assessment of credibility: legal testimony. Liverpool Law Review, 11, 181-199.

Siegler, R. S. (1991). Children's Thinking (2nd Ed.). New Jersey: Prentice-Hall, Inc.

Sink, F. (1988). A hierarchical model for evaluation of child sexual abuse. American Journal of Orthopsychiatry, 58, 129-135.

Snyder, M., & Swann, W. B., Jr. (1978). Behavioral confirmation in social interaction: From social perception to social reality. Journal of Experimental Social Psychology, 14, 148-162.

State v. French, 760 P. 2d 86 (Montana, 1988).

State v. Geyman, 729 P. 2d 475 (Montana, 1986).

State v. Kim, 645 P. 2d 1330 (Hawaii, 1982).

State v. Myers, 359 N.W. 2d 604 (Minnesota, 1984).

State v. Timperio, 528 N.E. 2d 594 (Ohio, 1987).

State v. Yates, 399 P. 2d 161 (Oregon, 1965).

Steller, M. (1989). Recent developments in statement analysis. In J. C. Yuille (Ed.), Credibility assessment (pp. 135-154). Dordrecht, The Netherlands: Kluwer.

Steller, M., & Köhnken, G. (1989). Criteria-based statement analysis. In D. C. Raskin (Ed.), Psychological methods in criminal investigations and evidence. New York: Springer.

Steller, M., Wellershaus, P., & Wolf, T. (1988, June). Empirical validation of criteria-based content analysis. Paper presented at the NATO Advanced Study Institute on Credibility Assessment in Maratea, Italy.

Szewczyk, H. (1973). Kriterien der beurteilung kindlicher zeugenaussagen. Probleme und Ergebuisse der Psychologie, 46, 47-66.

Szewczyk, H., & Littmann, E. (1982). Untersuchugen zur glaubweerdyheitsbeinteilung kinklicher zeugen. In A. Trankell (Ed.), Reconstructing the past: The role of psychologists in criminal trials (pp. 73-103). Stockholm: Norstedt & Soners.

Trankell, A. (1957). Zur Methodik der glaubwurdigkeitsuntersuchungen. Psychologie und Praxis, 1, 292-311.

Trankell, A. (1963). Vittnespsykologins arbetsmetoder. Stockholm: Bokforlager Liber.

Trankell, A. (1972). Reliability of Evidence. Stockholm: Beckmans.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185, 1123-1131.

U.S. v. Azure, 801 F. 2d 336 (8th Cir. 1986).

Uebersax, J. S. (1987). Diversity of decision-making models and the measurement of interrater agreement. Psychological Bulletin, 101, 140-146.

Undeutsch, U. (1967). Beurteilung der Glaubhaftigkeit von Aussagen. In U. Undeutsch (Ed.), Handbuch der Psychologie. Bd. II: Forensische Psycholgie. Goettingen: Hogrefe.

Undeutsch, U. (1982). Statement reality analysis. In A. Trankell (Ed.) Reconstructing the past: The role of psychologists in criminal trials (pp. 27-56). Stockholm: Norstedt & Soners.

Undeutsch, U. (1984). Courtroom evaluation of eye-witness testimony. International Review of Applied Psychology, 33, 51-67.

Undeutsch, U. (1989). The development of statement reality analysis. In J. C. Yuille (Ed.) Credibility Assessment (pp. 101-119). Dordrecht, The Netherlands: Kluwer.

Varendonck, J. (1911). Les temoignoges d'enfants dans un proces retentissant. Archives de psychologie, 11, 129-171.

Wakefield, H., & Underwager, R. (1988). Accusations of child sexual abuse. Springfield, IL: Charles C. Thomas.

Wakefield, H., & Underwager, R. (1991). Sexual abuse allegations in divorce and custody disputes. Behavioral Sciences & the Law, 9, 451-468.

Wells, G., & Loftus, E. (1991). Commentary: Is this child fabricating? Reactions to a new assessment technique. In J. Doris (Ed.), The suggestibility of children's recollections (pp. 168-171). Washington DC: American Psychological Association.

Wigmore, J. H. (1940). A treatise on the Anglo-American system of evidence in trials at common law, including the statutes and judicial decisions of all jurisdictions of the United States and Canada (3rd edition). Boston: Little, Brown, and Company.

Wyer, R. S. Jr. (1988). Social memory and social judgement. In R. R. Solomon, G. R. Goethals, C. M. Kelly, & B. R. Stephens (Eds.), Perspectives on Memory Research. New York: Springer-Verlag.

Yuille, J. C. (1988a). The systematic assessment of children's testimony. Canadian Psychology, 29, 247-262.

Yuille, J. C. (1988b, June). A simulation study of criteria-based content analysis. Paper presented at the NATO Advanced Study Institute on Credibility Assessment, Maratea, Italy.

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In L. Berkowitz (Ed.) Advances in experimental social psychology (Vol. 14, pp. 1-59). New York: Academic Press.

Table 1 Criteria-Based Content Analysis (Raskin & Esplin, 1991a)

General Characteristics

1. Logical Structure. Is the statement coherent? Is the content logical? Do the different segments fit together? (Note: Peculiar or unique details or unexpected complications do not diminish logical structure.)
2. Unstructured Production. Are descriptions unconstrained? Is the report somewhat unorganized? Are there digressions or spontaneous shifts of focus? Are some elements distributed throughout? (Note: This criterion requires that the account is logically consistent.)
3. Quantity of Details. Are there specific descriptions of place or time? Are persons, objects, and events specifically described? (Note: Repetitions do not count.)

Specific Contents

4. Contextual Embedding. Are events placed in spatial and temporal context? Is the action connected to other incidental events, such as routine daily occurrences?
5. Interactions. Are there reports of actions and reactions or conversations composed of a minimum of three elements involving at least the accused and the witness?
6. Reproduction of Speech. Is speech or conversation during the incident reported in its original form? (Note: Unfamiliar terms or quotes are especially strong indicators, even when attributed to only one participant.)
7. Unexpected Complications. Was there an unplanned interruption or an unexpected complication or difficulty during the sexual incident?
8. Unusual Details. Are there details of persons, objects, or events that are unusual, yet meaningful in this context? (Note: Unusual details must be realistic.)
9. Superfluous Details. Are peripheral details described in connection with the alleged sexual events that are not essential and do not contribute directly to the specific allegation? (Note: If a passage satisfies any of the specific criteria 4-18, it probably is not superfluous.)
10. Accurately Reported Details Misunderstood. Did the child correctly describe an object or event but interpret it incorrectly?
11. Related External Associations. Is there reference to a sexually-toned event or conversation of a sexual nature that is related in some way to the incident but is not part of the alleged sexual offenses?
12. Subjective Experience. Did the child describe feelings or thoughts experienced at the time of the incident? (Note: This criterion is not satisfied when the witness responds to a direct question, unless the answer goes beyond the question.)
13. Attribution of Accused's Mental State. Is there reference to the alleged perpetrator's feelings or thoughts during the incident? (Note: Descriptions of overt behavior do not qualify.)

Motivation-Related Contents

14. Spontaneous Corrections or Additions. Were corrections offered or information added to material previously provided in the statement? (Note: Responses to direct questions do not qualify.)
15. Admitting Lack of Memory or Knowledge. Did the child indicate lack of memory or knowledge of an aspect of the incident? (Note: In response to a direct question, the answer must go beyond "I don't know" or "I can't remember".)
16. Raising Doubts About One's Own Testimony. Did the child express concern that some part of the statement seems incorrect or unbelievable? (Note: Merely asserting that one is telling the truth does not qualify.)
17. Self-Deprecation. Did the child describe some aspect of his/her behavior related to the sexual incident as wrong or inappropriate?
18. Pardoning the Accused. Did the child make excuses for or fail to blame the alleged perpetrator, minimize the seriousness of the acts, or failed to add to the allegations when the opportunity occurred?

Table 2 Statement Validity Checklist (Raskin & Esplin, 1991a)

PSYCHOLOGICAL CHARACTERISTICS

1. Cognitive-Emotional Limitations - Are there indications that limited cognitive abilities, unwillingness to discuss the events, or discomfort during the interview interfered with obtaining adequate information from the interview process?
2. Language and Knowledge - Was the child's use of language and display of knowledge beyond the normal capacity for a child of that age and experience and beyond what the child may have learned from the incident?
3. Affect During the Interview - Did the child display inappropriate affect during the interview or was there an absence of affect that would be expected to accompany such a report by this child?
4. Suggestibility - Did the child demonstrate susceptibility to suggestion or ask questions during the interview to attempt to obtain clues as to what to say to the interviewer?

INTERVIEW CHARACTERISTICS

5. Interview Procedures - Was this interview inadequate according to principles and procedures of statement validity assessment? Did the interviewer introduce distractions, fail to establish rapport, inadequately attempt to elicit a free narrative, fail to use open questions and appropriate follow-up questions, or fail to attempt to resolve ambiguities and apparent inconsistencies? Were reasonable alternative hypotheses ignored?
6. Influence on Statement Contents - Was there leading or suggestive questioning, pressure, or coercion in any analyzed interview of the child? Were suggestive techniques or props employed in any interview?

MOTIVATIONAL FACTORS

7. Motives for Reporting - Does the child's relationship to the accused or other contextual variables (e.g., living arrangements or relationships among significant others) suggest possible motives for the child to make a false allegation?
8. Context of Disclosure - Are there questionable elements in the context of the original disclosure or report of the accusations? Are there important inconsistencies in the reports?
9. Influence by Others - Are there indications that others suggested, coached, pressured, or coerced the child to make a false report?

INVESTIGATIVE QUESTIONS

10. Lack of Realism - Are the described events unrealistic? Are there major elements in the statement that are contrary to the laws of nature?
11. Inconsistent Statements - Are there major elements in the statement (not peripheral details) that are inconsistent or contradicted by another statement made by this child or another witness?
12. Contradictory Evidence - Are there major elements in the statement that are contradicted by reliable physical evidence or other concrete evidence?
13. Characteristics of the Offense - Is the description of the alleged sexual offense lacking in the normal details and general characteristics of this type of offense against a child? Does the description contain important elements or general characteristics that are contrary to what has been established in the professional and investigative literature concerning such offenses?

Table 3 Undeutsch's Statement Reality Analysis Criteria (Undeutsch, 1989)

I. CRITERIA DERIVED FROM SINGLE STATEMENTS

A. GENERAL, FUNDAMENTAL CRITERIA
1. Anchoring (embodiment) in time and place
2. Concreteness (clarity and definiteness, specification, vividness)
3. Wealth of reported details
4. Originality (individual depiction, more than cliches, trite and stereotyped phrases)
5. Internal consistency
6. Mentioning of details, specific of the particular type of sex offense

B. SPECIAL MANIFESTATIONS OF THE AFOREMENTIONED CRITERIA
7. Reference to details that exceed the probable capacity of the witness
8. Reporting of subjective experiences
9. Mentioning of unexpected complications
10. Spontaneous corrections, specifications, complements
11. Self-disserving interspersions

C. NEGATIVE OR CONTROL CRITERIA
12. Lack of internal consistency
13. Lack of consistency with the laws of nature and science
14. Lack of external consistency (discrepancy with other incontrovertible facts)

II. CRITERIA DERIVED FROM SEQUENCES OF STATEMENTS

A. LACK OF PERSISTENCE (STABILITY, STEADFASTNESS)

B. PRIOR INCONSISTENT STATEMENT

Table 4 Percent Presence of Criteria in Confirmed and Doubtful Cases -- Esplin et al. (1988)

Criterion	Confirmed	Doubtful
1	100	55
2	95	15
3	100	55
4	100	35
5	100	30
6	70	0
7	70	0
8	95	0
9	100	5
10	5	5
11	90	0
12	90	30
13	40	0
14	100	10
15	75	35
16	10	0
17	25	0
18	55	5
19	100	30

Table 5 Inter-rater Reliability of the Ratings of Criteria (categorized by Maxwell's RE) -- Anson et al. (1993)

Group/criterion	Maxwell's RE	Cohen's kappa	Proportion Agreement	Average Presence
Group 1: Adequate				
Pardoning accused (18)	1.00	1.00	1.00	.09
Doubting testimony (16)	1.00	N/A	1.00	.00
Accused's state (1)	.83	.70	.91	.17
Misunderstood details (10)	.83	.47	.91	.09
Self-deprecation (17)	.74	.51	.87	.15
Speech (6)	.65	.64	.83	.61
Details (3)	.65	.55	.83	.74
Logical structure (1)	.65	.01	.83	.91
Complications (7)	.56	.40	.78	.24
Group 2: Marginal				
Superfluous details (9)	.48	.47	.74	.57
Contextual embedding (4)	.48	.32	.74	.74
Unusual details (8)	.39	.06	.70	.19
Corrections (14)	.39	.04	.70	.20
Group 3: Inadequate				
Lack of memory (15)	.22	.16	.61	.37
External associations (11)	.22	.03	.61	.28
Interactions (5)	.13	.14	.57	.48
Subjective experience (12)	.13	.09	.57	.61
Unstructured production (2)	.13	.01	.57	.70
Details of offense (19)	-.22	-.30	.39	.65

Table 6 Criterion Use for True and False Senders -- scores are between 0 - 12 (Landry & Brigham, 1992)

Criterion	True	False	T-test	Frequency
Logical Structure (1)	6.79	7.71	-3.84***	86%
Quantity of Details (3)	7.80	7.06	2.57*	84%
Contextual Embedding (4)	5.82	4.52	4.14***	66%
Interactions (5)	6.18	5.48	2.37*	71%
Reproduction of Speech (6)	3.46	1.37	9.80***	30%
Complications (7)	3.15	3.37	.81	43%
Unusual Details (8)	5.20	4.58	2.18*	58%
Superfluous Details (9)	4.91	3.97	3.70***	54%
Subjective Experience (12)	8.31	7.18	5.74***	85%
Accused's Mental State (13)	2.28	4.09	-6.31***	39%
Spontaneous Corrections (14)	1.34	.61	4.47***	13%
Admits Lack of Memory (15)	.46	.15	3.00**	4%
Doubt Own Testimony (16)	1.11	.32	5.43***	9%
Self Deprecation (17)	.89	1.06	.97	13%
Total of 14 Criteria	57.69	51.49	4.33***	

* $p < .05$, ** $p < .01$, *** $p < .001$.

Table 7 Summary of the Usefulness of CBCA Criteria. An "X" signifies the criterion was useful to some extent in distinguishing between true and false statements, an "o" signifies that the criterion was not, a blank space signifies that the researchers did not assess that criterion. See the particular study for standards used to determine usefulness.

Study	Criteria																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Field Studies:																			
Littmann & Szewczyk (1983)	o	o	o	o	o	o	o	o	o	o	o	o	o	o	X	o	o	o	o
Esplin et al. (1988)	X	X	X	X	X	X	X	X	X	o	X	X	o	X	X	o	o	X	X
Boychuck (1991)	X	X	X	X	X	X	X	X	o	o	X	X	o	X	o	o	o	o	X
Anson et al. (1993)	X	X	X	X	o	X	o	o	X	o	o	X	o	o	o	o	o	o	X
Experimental Studies:																			
Köhnken & Wegener (1982)		o	X																
Steller et al. (1988)	X	o	X	X	X	X	X	X	X	X	X	X	o	o	o		o	o	
Landry & Brigham (1992)	o		X	X	X	X	o	X	X			X	o	X	X	X	o		
Köhnken et al. (1993)	o	X	X				o	o	o	o			o	o	o	o		o	

Table 8 Percentages of Statements Correctly Classified as True or False in Experimental Studies

Study:	True	False
Yuille (1988b)	91%	74%
Steller et al. (1988)	78%	62%
Landry & Brigham (1992)	75%	35%